



**icbibi**

THE INNOVATION CENTER FOR BIOMEDICAL INFORMATICS

# **G-DOC *Plus*** **Precision medicine use case**

*Innovation Center for Biomedical Informatics*  
*Georgetown University*

# Background

- Explore somatic and germline variants in cancer samples
- Inactivation of tumor suppressor genes could be caused by accumulation of mutations or, loss of heterozygosity (LOH), causing progression of cancer
  - one such location is Chromosome 8,
  - abnormalities in Chr 8 reported in BRC, CRC
- Example query: ***which genes are affected by novel deletions in chromosome 8 with potential impact on protein function, and what major pathways might these genes be involved in?***

# Overview

- Register
- Login
- Navigation
- Precision medicine workflow
  - Variant search

# First time user



The Innovation Center for Biomedical Informatics (ICBI)  
Lombardi Comprehensive Cancer Center

Thu Jan 22, 2015

Log In

[register now](#) | [forgot password](#)



Welcome to GDOC Plus Beta

Precision Medicine

Registration with Georgetown Net ID (or any other email). You will get an email with a link that you need to click to confirm registration

Population genetics

## Understanding Data in G-DOC Plus

It all begins with a study...

All data in G-DOC Plus derives from studies on topics such as breast cancer, wound healing, or even 1,000 Genomes. Each study may contain clinical and/or biospecimen data. Below is an overview of studies by topic.

\* private studies, ones which are uploaded and marked private, are not counted here

### News

October 02, 2014: ICBI Symposium 2014

[\[read\]](#)

May 02, 2014: Featured in Frontiers' Top 10 2013 Most viewed Genetics Research articles

[\[read\]](#)

March 12, 2014: AAAS Big Data Blog [\[read\]](#)

# Login



The Innovation Center for Biomedical Informatics (ICBI)  
Lombardi Comprehensive Cancer Center

Thu Jan 22, 2015

Log In

[register now](#) | [forgot password](#)



## Welcome to GDOC Plus Beta!

The Georgetown Database of Cancer Plus other diseases (G-DOC Plus) is a precision medicine platform containing molecular and clinical data from thousands of patients and cell lines, along with tools for analysis and data visualization. The platform enables the integrative analysis of multiple data types to understand disease

**Precision Medicine**

**Translational research**

**Population genetics**

## Understanding Data in G-DOC Plus

It all begins with a study...

All data in G-DOC Plus derives from studies on topics such as breast cancer, wound healing, or even 1,000 Genomes. Each study may contain clinical and/or biospecimen data. Below is an overview of studies by topic.

\* private studies, ones which are uploaded and marked private, are not counted here

## News

October 02, 2014: ICBI Symposium 2014

[\[read\]](#)

May 02, 2014: Featured in Frontiers' Top 10 2013 Most viewed Genetics Research articles

[\[read\]](#)

March 12, 2014: AAAS Big Data Blog [\[read\]](#)

# G-DOC Plus Launch Pad!

Welcome! The G-DOC Plus Launch Pad is your one-stop resource for learning more about G-DOC and getting started on the platform.



Studies



Lists



**It All Starts Here!**



G-DOC has over seventy studies, We know this can be overwhelming! Let us guide you to choose the study that is relevant for your research.

Let's Go! >



Groups



Notifications **0**

# Selections

- "What's your area of interest" : Precision medicine - it should go to the workflow page.
- Select "data collection" : BREAST CANCER
- Select a study : BRC\_CG\_XXXX\_01 (Note: If you click on "More", you will be able to see the complete description of the study)
- Select the tool: Variant Search

## What's your area of interest?

G-DOC Plus has three overlapping entry points for the user based on their interests. Choose your area of interest to launch the workflow.

**Precision Medicine**  
Patients' molecular diagnostics and clinical data.

**Translational Research**  
Analytic tools and workflows to enable discovery.

**Population Genetics**  
Race-based, genomic reporting and comparison.

## Precision Medicine

What type of data collection do you want to analyze?

**BREAST CANCER**

2 studies	93 samples	4 biospecimen
-----------	------------	---------------

## Precision Medicine

Study Selected! BREAST CANCER → BRC\_CG\_XXXX\_01

Based upon the study you picked, here is a list of tools you can use:

Search

- Variant Search

## Precision Medicine

Great! What breast cancer study would you like

**BRC\_CG\_XXXX\_01**

**Title:** Complete Genomics Breast Cancer dataset

**Data Type Details:** CLINIC,WGS

**Abstract:** Sample dataset from Complete Genomics Cancer Sequencing Service. It contains matched tumor and normal cell line sequence data for two patients with breast cancer. Collections were drawn from ATCC. Ruminant were sequenced in an average

4 samples	4 biospecimen
-----------	---------------

More>>

**BRC\_IMAGING\_TCIA\_01**

**Title:** Breast-Diagnosis imaging collection from TCIA

**Data Type Details:** IMAGING,CLINIC

**Abstract:** The Breast-Diagnosis collection from TCIA contains cases that are high-risk normals, DCIS, fibroids and lobular carcinomas. Each case has 3 or more distinct MR pulse sequences from a Philips 1.5 T (fused sequences are labeled T2, STR and

89 samples	0 biospecimen
------------	---------------

More>>

**Tip:** Click "More" to see complete description of study. Click "Select study"

- Make the following selections:
  - **Chromosome:** *chromosome 8*, **Functional Location:** *Coding sequence*, **Exonic Function:** *Non Synonymous -> Deletion*, **Variant inclusion :** *Novel*

Data set: WGS from breast cancer cell line

Data stored in Amazon cloud



## Search Sequence Variations

Current Study: BRC\_CG\_XXXX\_01 [change study?](#)

**Genes**

**Chromosome**

**Functional Location**

**Exonic Function**

Non-Synonymous

- Loss of stop
- Premature stop
- Insertion
- Deletion
- Frameshift
- Substitution
- Possible splice variant
- Possible 5' splice variant
- Other

**Variant Inclusions**

Known

Novel

**Sample IDs (max 2)**

Chromosome	Start	Referenc	Alleles	XRefs	Gene	Funcitonal Lo	Exonic Function	REF_FREQ	ALT_FREQ	MISSING_
chr8	10480173	GCGGC	G	.	RP1L1	Coding seque	Deletion Frameshift	0.625	0.125	0.25
chr8	23186052	ACCG	A	.	LOXL2	Coding seque	Deletion	0.875	0.125	
chr8	77765302	CCTC	C	.	ZFHX4	Coding seque	Deletion	0.625	0.25	0.125
chr8	103573010	CTGCAA	C	.	ODF1	Coding seque	Deletion	0.125	0.125	0.75
chr8	145113143	TCCC	T	.	OPLAH	Coding seque	Deletion Possible 5' splice	0.875	0.125	
chr8	145625537	CC	G	.	CPSF1 MIR1234	Coding seque Non-coding L Intronic	Deletion Frameshift	0.25	0.125	0.625
chr8	145756159	GC	G	.	ARHGAP	Coding seque	Deletion Frameshift	0.875	0.125	

Save gene list Export results
Page 1 of 1 50 View 1 - 7 of 7

Save gene list. Name: BRC-8-CDS-Del\_Novel

- Once the gene list is saved, click on **Lists on the Top panel**. You will see the gene list grouped under the study name (BRC\_CG\_XXXX\_01)
- On the right side there will be several icons which indicate further downstream analysis. Click on the green icon which indicates **Enrich Gene List** analysis

## Lists

Below are all G-DOC Plus saved Lists. You can view, modify, upload, export and share Lists with other groups.

You can use the filter below to search for a specific list. You can also use the tool Panel on the right to manipulate data in the saved lists.

Filter

Check all for deletion Delete List(s) Upload

Tags	Study	Lists
	AD_BLALOCK_2011_01	
	ALL_NORDLUND_2013_01	
	BRC_BUFFA_2011_01	
	BRC_CG_XXXX_01	
gene	BRC_CG_XXXX_01	<input type="checkbox"/> <b>BRC-B-CDS-Del_N</b> (8 items) <span style="float: right;">3:49 2/10/2015</span> <div style="border: 1px solid gray; padding: 5px; margin-top: 5px;">           Studies: BRC_CG_XXXX_01            Tags: gene         </div>

# Pathway enrichment using Reactome

## Pathway Enrichment Results

Pathway	p-value	Overlapping Genes
Inactivation of Cdc42 and Rac	$5.483 \times 10^{-3}$	ARHGAP39
Glutathione synthesis and recycling	$6.699 \times 10^{-3}$	OPLAH
Processing of Intronless Pre-mRNAs	$8.521 \times 10^{-3}$	CPSF1
Post-Elongation Processing of Intronless pre-mRNA	$1.397 \times 10^{-2}$	CPSF1
Processing of Capped Intronless Pre-mRNA	$1.397 \times 10^{-2}$	CPSF1
Glutathione conjugation	$1.518 \times 10^{-2}$	OPLAH
Signaling by Robo receptor	$1.941 \times 10^{-2}$	ARHGAP39
Post-Elongation Processing of Intron-Containing pre-mRNA	$2.061 \times 10^{-2}$	CPSF1
mRNA 3-end processing	$2.061 \times 10^{-2}$	CPSF1
Transport of Mature mRNA Derived from an Intronless Transcript	$2.121 \times 10^{-2}$	CPSF1
Transport of Mature mRNAs Derived from Intronless Transcripts	$2.181 \times 10^{-2}$	CPSF1
Cleavage of Growing Transcript in the Termination Region	$2.602 \times 10^{-2}$	CPSF1
Post-Elongation Processing of the Transcript	$2.602 \times 10^{-2}$	CPSF1
RNA Polymerase II Transcription Termination	$2.602 \times 10^{-2}$	CPSF1
Rho GTPase cycle	$2.662 \times 10^{-2}$	ARHGAP39
Signaling by Rho GTPases	$2.662 \times 10^{-2}$	ARHGAP39
Gene Expression	$3.140 \times 10^{-2}$	CPSF1

Export results

Page 1 of 1 50

View 1 - 27 of 27

# Discussion

- Pathways enriched - related to processing of pre-mRNA and mature mRNA,
  - are known to potentially alter gene expression and in-turn turn tumor suppressor genes off
- The top pathway “Inactivation of CD42 and RAC”
  - shows that when such RNA processing pathways are turned negative, it causes other tumor pathways to go out of control.

# Application

- Case study shown on cell line data - represents individual tumors
  - can be applied on other patient datasets
  - allows in-depth analysis of germline mutations for individual patients based on NGS data.
- The variant search tool also allows exploration of somatic variants (variants present in tumor, but not in matched normal sample)

- To **view the list of genes** in the saved gene list, once again go to **Lists** on the top panel.
- Select the middle icon on the right side which indicates **Export list**. You will now see a text file downloaded to your system



- You can open the exported file to view the list of genes



RP1L1  
CPSF1  
ODF1  
LOXL2  
ZFHX4  
OPLAH  
MIR1234  
ARHGAP39

# Appendix

- **Functional Location:** Physical location of Variants on the chromosome.
  - Functional location filter options: coding sequence, intronic, downstream of gene, upstream of gene, 5' UTR, 3' UTR, non-coding UTR, intergenic
- **Coding gene:** Within a gene which codes for a protein. Within this region are search terms for Intron, downstream of the specified gene, upstream of the specified gene, in the 5' untranslated region, 3' untranslated region.
- **Non-coding region:** Region that does not cover coding genes, it's flanking regions and UTR.
- **Intergenic:** Excludes Coding regions and non-coding UTRs.
- **Non-Synonymous changes**
  - **Exonic Functions:** Includes regions covered only exons of a gene coding for a protein where variation has resulted in non-synonymous changes.
  - Exonic function filter options: Loss of stop (structurally impacted protein), Premature stop (truncated protein), Insertion, Deletion, Frameshift, Substitution, Possible splice variant, Possible 5' splice variant, other

# General tips

- G-DOC *Plus* works best if you don't use the **back** button in the web browser repeatedly.
- Once you select a study, most tools will be easily available from the the top menu bar inside G-DOC *Plus*.
- The Pathway enrichment and the Lists tool may sometimes take a few seconds longer to execute than other tools (since they are directly connecting to the server every time). Your patience is highly appreciated.

# Clearing cache

- If the G-DOC web page does not respond after several seconds, try:
  - refreshing the page.
  - Log out and log back in, and try again
  - If the above two do not work, its possible that your web browser cache may need to be cleared
    - For Google chrome, go to **Settings** -> **Show Advanced Settings** -> Under “Privacy”, select **Clear Browsing data**
    - For Mozilla Firefox, go to **Preferences** -> **Advanced** -> **Network** -> Under “Cached Web Content” -> **Clear now**

